

SCIENTIFIC REPORTS



Corrected: Author Correction

OPEN

Illumina sequencing of clinical samples for virus detection in a public health laboratory

Bixing Huang¹, Amy Jennison², David Whiley^{3,4}, Jamie McMahon¹, Glen Hewitson¹, Rikki Graham², Amanda De Jong¹ & David Warrilow¹

High-throughput sequencing (HTS) provides the opportunity, once a diagnostic result is obtained, to extract additional information from a virus-containing sample. Hence, it offers advantages over established quantitative amplification technology, such as quantitative PCR, particularly in a public health environment. At this early stage of its clinical application, there have been limited studies comparing HTS performance to that of the more established quantitative PCR technology for direct detection of viruses. In this pilot-scale study, we tested HTS with a range of viruses and sample types routinely encountered in a public health virology laboratory. In comparison with quantitative PCR, our HTS method was able to sensitively (92%) detect all viruses in any sample type with the exception of certain tissues. Moreover, sufficient nucleotide sequence information was obtained to enable genotyping of strains detected, thus providing additional useful epidemiological information. While HTS sensitivity may not yet match that of PCR, the added value through enhanced epidemiological data has considerable potential to enable real-time surveillance of circulating strains so as to facilitate rapid and appropriate response to outbreaks and virus zoonotic spillover events.

The technology broadly referred to as “next generation sequencing” or high-throughput sequencing (HTS) uses various chemistries to obtain large amounts of data about the nucleic acid composition of a sample. The technology is used in multiple areas from the biological sciences to health. For the latter, it has been applied to the diagnosis of inheritable diseases, cancer, and infectious diseases. HTS potentially enables everything in the sample to be sequenced. This unbiased aspect of HTS has facilitated rapid large-scale virus discovery, as well as the detection of unusual or novel viral agents associated with human disease^{1–11}. As well as exploring cases of unknown disease etiology, HTS offers the prospect of combining diagnostics with sequence analyses to reveal clinically relevant information which can be used to enhance public health disease interventions and disease surveillance. In virology, this includes molecular epidemiology and genotyping for outbreak investigations, exclusion testing for other non-viral pathological agents and the identification of antiviral resistance mutations and virulence determinants.

Whilst HTS has become widely adopted in the infectious disease area over the last 10 years, and the technology and supporting bioinformatics substantially improved, it is still considered to be an immature technology. HTS's strengths complement the polymerase chain reaction (PCR), now considered the primary diagnostic workhorse. However, the frequency with which changes are made to the basic chemistry and the associated hardware and software necessitate ongoing reassessment of its use. Interestingly, whilst there are many examples of its application to microbiological metagenomics and detection of novel agents, there are few examples in the literature comparing its diagnostic efficacy to more conventional assays for direct detection of viruses in clinical samples. Of the studies conducted so far, the most thorough comparisons have involved comparing HTS with PCR for direct detection of viruses in respiratory specimens^{12–16}. In these studies, correlations were found between viral load or quantitative PCR cycle threshold number (C_T ; the cycle where the fluorescent signal from the product amplification starts to exceed a baseline value and being a value that is inversely proportional to viral load), and numbers of normalized reads matching the target. Moreover, all the studies except one¹⁶ observed levels of

¹Public Health Virology Laboratory, Queensland Health Forensic and Scientific Services, PO Box 594, Archerfield, Queensland, 4108, Australia. ²Public Health Microbiology Laboratory, Queensland Health Forensic and Scientific Services, PO Box 594, Archerfield, Queensland, 4108, Australia. ³Microbiology Division, Pathology Queensland Central Laboratory, Brisbane, Queensland, 4029, Australia. ⁴Faculty of Medicine, University of Queensland Centre for Clinical Research, The University of Queensland, Brisbane, Queensland, 4029, Australia. Correspondence and requests for materials should be addressed to D. Warrilow (email: David.Warrilow@health.qld.gov.au)

HTS sensitivity equivalent to, or approximating, that of PCR. Another study investigating hand, foot and mouth disease virus detection in clinical specimens over a range of viral loads also observed comparable sensitivity to reverse-transcription quantitative PCR (RT-qPCR)¹⁷. A further study using samples spiked with various different viral agents confirmed the importance of viral load, with virus detected by HTS in all samples except those with PCR C_T values more than or equal to 37 cycles¹⁸. Overall HTS technology is looking promising for direct viral detection, however given the limited number of these studies to date, there is a need to examine the technology over a wider range of viruses and sample types.

We are a routine public health virology laboratory, and in this study we sought to explore the utility of HTS methodology in our setting. Briefly, the ability of HTS to directly detect viruses was compared with established PCR assays for viruses and sample types commonly encountered in the viral diagnostic laboratory. The viruses tested included those with single-stranded RNA, double-stranded RNA and double-stranded DNA genomes; and genome sizes from ~10 Kb to ~200 Kb. The clinical samples types tested included serum, urine, feces, swabs of various sites, and brain tissue. In addition to measuring clinical sensitivity, a key aim was to examine the additional nucleotide sequencing data that could be obtained for downstream phylogenetic analysis, so as to assist with outbreak investigations. Molecular epidemiology is an integral part of our public health investigative processes, and so HTS directly on clinical material has the potential to help target effective responses.

Materials and Methods

Clinical samples for sequencing. Clinical samples ($n = 52$) of various kinds, including serum ($n = 18$), urine ($n = 2$), feces ($n = 8$), nappy swabs ($n = 2$), skin swabs ($n = 5$), respiratory swabs ($n = 14$), and tissue ($n = 3$), were chosen on the basis of previously being identified as containing a virus by either RT-qPCR (RNA viruses, $n = 47$) or qPCR (DNA viruses, $n = 5$)¹⁹. These are detailed (virus, sample type, numbers, and C_T values) in Table S1 (supplementary file) and comprised a mix of viruses commonly encountered by the laboratory (including influenza, measles, mumps, arboviruses, molluscum contagiosum, rotavirus and noroviruses) and having a spectrum of low to high viral based on previously determined C_T values. The samples were submitted to the Public Health Virology Laboratory for routine viral testing over the period January 2015–March 2017. Further details of the PCR assays used for routine diagnostics are provided in Table S2. Please note that sequence data for a chikungunya virus and 7 norovirus samples (indicated in Table S1) were previously published; however, here we further analyzed the data for its diagnostic relevance. All work handling clinical material was done at biocontainment level 3.

Limit of detection estimation using bovine viral diarrhea virus (BVDV). A single-stranded and non-segmented RNA virus, an isolate of BVDV (strain MD-73), was used as a source of template for positive control libraries and for limit of detection experiments. The isolate was grown in Madin-Darby bovine kidney cells (MDBK) by serial passage at 37 °C in growth medium (Opti-MEM with 3% fetal calf serum). RNA was extracted from tissue culture fluid as for clinical samples. BVDV genomic RNA was serially diluted and the C_T value determined for each dilution by RT-qPCR. Dilutions with C_T values of 19, 25, 30, 33 and 37 were used in the construction of Illumina Nextera cDNA libraries as described for clinical samples above.

High-throughput virus genome sequencing. Nucleic acid was extracted manually using the QIAamp Viral RNA extraction kit (Qiagen) as per the manufacturer's instructions, omitting carrier RNA from the sample. Depending on whether the virus was RNA or DNA (based on the original PCR results), either an RNA or DNA library was prepared. RNA libraries were prepared as described previously²⁰ using an equivalent amount of extract as for qPCR (5 μ L of 60 μ L total). Briefly, host DNA was removed by DNase treatment using heat-labile Heat & Run (ArcticZymes) followed by inactivation at 80 °C for 5 min. First-strand cDNA was generated using Protoscript II kit (New England Biolabs) with the supplied random primer mix followed by second strand DNA synthesis using a cocktail of enzymes including *Escherichia coli* DNA ligase, DNA polymerase I, and RNase H (New England Biolabs). Libraries were constructed with the Nextera XT library kit (Illumina) using barcoded primers. A negative control library prepared from fetal calf serum and a positive control library prepared from a BVDV RNA spiked sample (diluted to C_T 25), were also included. DNA libraries were prepared as for RNA libraries, omitting a DNase pre-treatment and cDNA synthesis steps. Library quality was assessed on a TapeStation prior to sequencing. Libraries passing the quality check were sequenced with a v2 mid-output kit on an Illumina NextSeq 500 machine with a minimum of 4.6 million reads (2×150 nt paired).

Genome assembly. Bioinformatics was performed using a method developed in-house. Draft genomes were assembled by mapping reads to closely-related complete reference sequences obtained manually from GenBank (<https://www.ncbi.nlm.nih.gov/>) using the inbuilt Geneious Mapper at the lowest sensitivity settings with no fine tuning on Geneious R11 software²¹. The proportion of reads corresponding to the viral target was determined by calculating the normalized reads (matches per million total reads, per kilobase total genome size). Genome sizes were obtained from ViralZone <https://viralzone.expasy.org/>. Normalized reads were presented as boxplots, and were constructed using the online version of BoxPlotR (<http://shiny.chemgrid.org/boxplotr/>).

Phylogenetic analysis. Molluscum contagiosum genome alignments were performed with Mauve²² as it can process genomes on this scale (~200 Kb). Measles and mumps virus genome alignments were performed with MAFFT v7.388²³ (algorithm set to "Auto"; scoring matrix 200PAM/ $k = 2$; a gap open penalty of 1.53; and an offset value of 0.123). Phylogenetic trees were constructed from the alignments with FastTree v2.1.5²⁴ as part of the Geneious R11 package (Jukes-Cantor model, 20 categories of sites and optimized for Gamma20). Branch support values were calculated using the Shimodaira-Hasegawa test and are shown as a percentage.

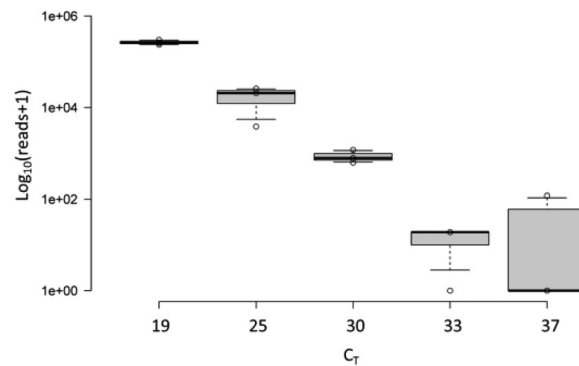


Figure 1. An estimation of the limit of detection for BVDV detection. (A) logarithmic representation of the matching reads in relation to the C_T value is shown for each dilution of BVDV. Individual data points are given ($^{\circ}$); the boxplots give the 1st and 3rd quartile, and whiskers 5th and 95th percentile.

Ethical approval. This work was approved by the Queensland Health Forensic and Scientific Services Human Ethics Committee in accordance with the Australian NHMRC National Statement on Ethical Conduct in Human Research 2007. The committee decided that the need for patient consent was waived for this study. All samples were submitted for routine diagnostics and had been de-identified. It should be noted that a condition of the ethical approval was that we were restricted to detecting only viruses known to be present in the sample and were prevented from attempting to use HTS to detect viruses other than those specified in the original pathology request. Hence, we were unable to determine the specificity of the HTS method.

Results

Limit of detection using BVDV. BVDV RNA was used as a surrogate to enable an estimation of the approximate limit of detection of HTS for clinical samples. Libraries were constructed in triplicate from samples spiked with serially diluted BVDV. A comparison of the C_T values with the logarithm of the normalized reads matching the target (Fig. 1) revealed a correlation ($r = -0.97$, Pearson's correlation co-efficient, $P < 0.05$; $R^2 = 0.95$). HTS detection was reproducible except for at the lowest BVDV RNA concentrations where no reads were detected in some replicates; virus was detected by HTS in two of three replicates at a C_T of 33 and only one of three replicates at a C_T of 37. Our accepted cut-off for detection is normally a C_T of <40 . Assuming product doubling every cycle, these data therefore show that the reliable limit of detection of HTS is approximately 1–2 logs less than that of qPCR; albeit, the C_T being only an approximation, not an absolute measure, of viral load.

Detection of virus in clinical samples using HTS. Libraries were prepared from a selection of 52 specimens shown to contain virus by qPCR. HTS detected the respective virus in 48 (92%) of these samples. With the exception of brain tissue ($n = 3$), reads matching the target and normalized to genome size (Fig. 2A) were inversely correlated with the C_T value ($r = -0.76$, Pearson's correlation co-efficient, $P < 0.05$; $R^2 = 0.59$; Note that samples with no matches to the target were excluded from calculation of the co-efficient). In the case of brain tissue, the number of matching reads was lower than that expected from the C_T value. It is likely that residual DNA from the sample competed with the viral target cDNA for incorporation into the library, and suggests the method may need further optimization for sequencing tissue.

A boxplot of normalized reads for each sample type is provided in Fig. 2B. Fecal samples (feces and nappy swab) had higher median values than other sample types ($P < 0.01$, Mann-Whitney test). This indicated higher viral loads in these samples, and was most likely due the fact that they primarily contained norovirus, which is known to have a high viral load in feces. With regard to genome type, only the dsRNA had noticeably higher normalized reads but was not statistically significant (Fig. 2C). There was no significant difference between RNA and DNA genomes ($P = 0.84$, Mann-Whitney test), plus-sense ssRNA and minus-sense ssRNA ($P = 0.10$, Mann-Whitney test); however, this was on the basis of 5 DNA virus samples and ideally more should be tested. Nor was there a significant difference between monopartite and multipartite genomes (Fig. 2D), both RNA and DNA included ($P = 0.12$, Mann-Whitney test). Hence, with the exception perhaps of tissue, the method is amenable to all sample and viral genome types.

Genome coverage was defined as the percentage of genome for which a contiguous sequence was obtained. The median genome coverage (Fig. 2E) for all viruses was 89% (20, 100; 1st and 3rd quartiles, respectively). Genome coverage for all viruses is shown graphically (Fig. 3). Interestingly, where there were sufficient reads obtained, each virus species seemed to have a distinctive and remarkably consistent read profile. This suggests one, or a combination, of two possibilities. Firstly, primer binding for reverse transcription is dependent on the structure of the RNA genome. This explanation is supported by the seemingly flatter read profile of the molluscum contagiosum samples; however, the much larger DNA genome size of this virus could make this deceptive. The uneven profile of the rotavirus samples with dsRNA genome, which should have minimal genome structure, argues against this. Alternatively, aspects of the library creation process, such as the transposase recombination step, may not be entirely random, leading to biases in those sequences incorporated into the library.

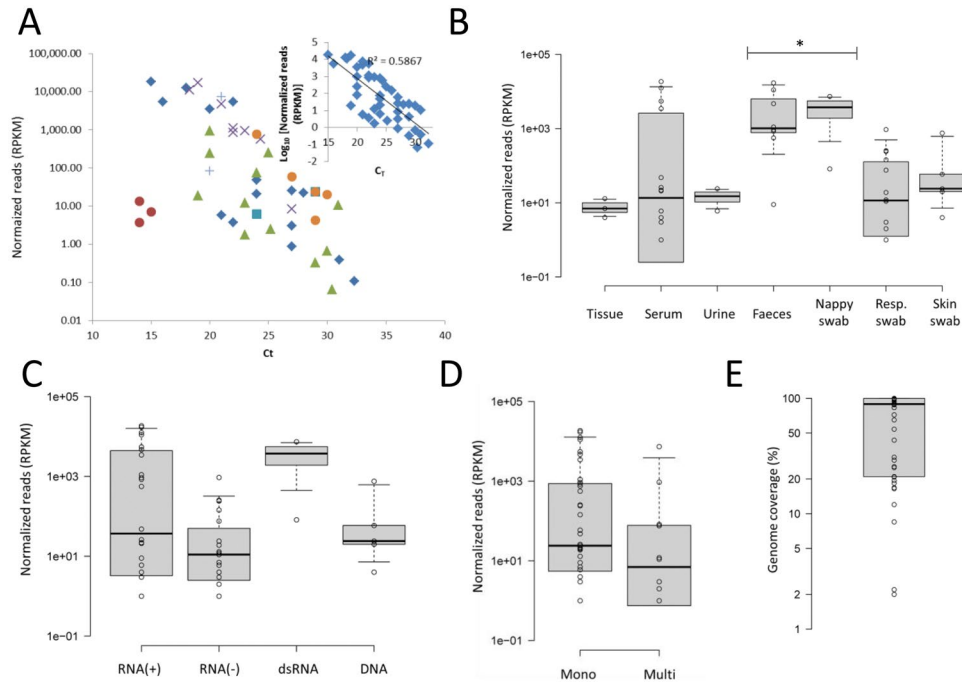


Figure 2. HTS assay performance with virus genomes and sample types. (A) Normalized reads (matching reads per million total reads per genome size, RPKM) are given for each sample type (tissue, ●; serum, ◆; urine, ■; faeces, ×; nappy swab, +; respiratory swab, ▲; and skin swab, ○). Data was used to calculate a correlation (inset). (B) Sample types data represented as boxplots. Samples that were significantly different to the other types are indicated (* $P < 0.01$; Mann-Whitney). Normalized reads comparing (C) genome nucleic acid and (D) partite structure. (E) Coverage obtained shown as a percentage of total genome size. Individual data points are shown (○); the boxplots give the 1st and 3rd quartile, and whiskers 5th and 95th percentile.

Genotyping for disease surveillance and outbreak investigations. The epidemiology of some samples included in this study, a case of importation of chikungunya virus and 7 noroviruses cases associated with outbreaks in aged care facilities, were published previously^{20,25}. Phylogenetic-based genotyping was attempted for an additional 4 of the 48 samples providing genome sequence data, and these samples are indicated in Table S1. A nucleotide alignment was performed on the measles genome sequence obtained from a 21 year-old female. Branch support confidently placed the genome in the genotype D8 (Fig. 4A). This was consistent with previous observations that the D8 genotype was the second highest prevalent genotype responsible for cases reported in New South Wales, Australia, over the period 2002–2011²⁶. Identical methodology was used to type two cases of mumps virus infection. One sample was from a 32 year-old female and the other from a 49 year-old male. The latter showed classic parotitis on the left side. These were both placed within the G genotype (Fig. 4B), but from different circulating strains due to the genetic distance separating the two sequences. We are not aware of this genotype being reported in Australia before; however, it has been widely reported in Western countries in recent years²⁷. Finally, a molluscum contagiosum genome from a 63 year-old male (Fig. 4C) was able to be genotyped as subtype 1, the most prevalent subtype²⁸. These examples, and those from the previously published cases, demonstrate the value of application of this technology to public health.

Discussion

There is a clear need to determine how unbiased HTS performs with different viruses and sample types. Here, we conducted a pilot study using the most commonly used HTS platform (Illumina). In this setting, the detection of associated unusual and novel viral agents is a real possibility. Hence, we have used minimal sample and data manipulation (i.e. no filtration) to reduce the possibility of bias. Our main interest is the direct detection of RNA viruses frequently encountered by our public health laboratory. In doing so, we examined HTS capability to detect viruses of various genome sizes and types, and in a variety of commonly encountered clinical samples and viral loads. This is a major potential advantage of using HTS over conventional methods as a diagnostic tool. In comparison with qPCR, the method was found to detect 92% of the positive samples. Our results compare favorably with the other similar studies described previously^{12–16}, and extends the comparison of HTS with PCR to clinical material other than previously published studies on respiratory samples.

Similarly to the previous comparative studies, HTS normalized reads from the various samples (tissue excepted) were inversely correlated to C_T values, and were found to approximate the line of best fit. This suggested that the amount of non-viral nucleic acids competing for library incorporation in these samples was similar. This observation was surprising, as it was expected that different sample types might vary in the relative amount of nucleic acid from the host and contaminating bacteria. It suggested that any of these samples may be used for

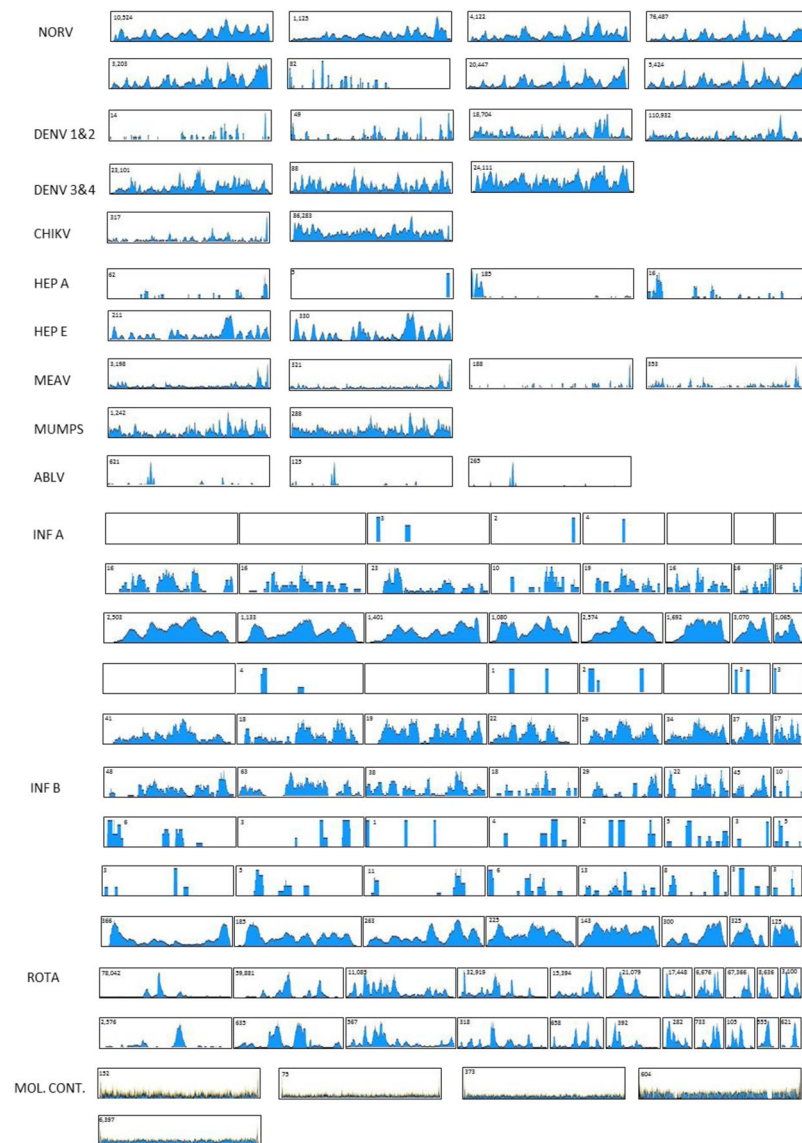


Figure 3. HTS contiguous assemblies. Assemblies are shown for each of the positive samples. Viruses shown are norovirus (NORV), dengue viruses 1 and 2 (DEN 1&2), dengue viruses 3 and 4 (DEN 3&4), chikungunya virus, (CHIKV), hepatitis A (HEP A), hepatitis E (HEP E), measles virus (MEAV), mumps virus (MUMPS), Australian bat lyssavirus (ABLV), influenza A virus (INF A), influenza B virus (INF B), rotavirus (ROTA), molluscum contagiosum (MOL. CONT.). Each box presents a single contiguous sequence (contig). Multipartite genome assemblies are shown in size relative to their other genome segments. Other contigs are all presented as the same size regardless of the genome length. The maximum read depth for each contig is given in the top left-hand corner of each assembly diagram.

HTS provided that viral load is high enough, and the number of sequencing reads is sufficient. Regarding the latter, the result also indicates that the minimum number of reads to be collected to be confident of virus detection could be similar for most sample types. We used approximately 5 million reads as a minimum, however further assessments of this may be warranted given the relatively low number of individual sample types tested.

Importantly, from a public health investigation perspective, we were able to use HTS to identify both previously identified and unidentified genotypes directly in clinical samples. This included information on an imported case of chikungunya virus infection; and genotype information of strains related to norovirus outbreaks in aged care facilities, measles, mumps and molluscum contagiosum. Notably, this ability to obtain significant epidemiologically useful sequence information directly from clinical samples represents a key potential use of HTS to public health laboratories with access to this technology. That is, even if HTS cannot ultimately match the diagnostic sensitivity of PCR, it could otherwise prove to be an invaluable reflex test to PCR where targeted PCR methodology has been unable to identify a causative agent, for surveillance purposes, and to provide enhanced epidemiological information during an outbreak. Certainly, our intention is not to suggest that HTS is a replacement for PCR.

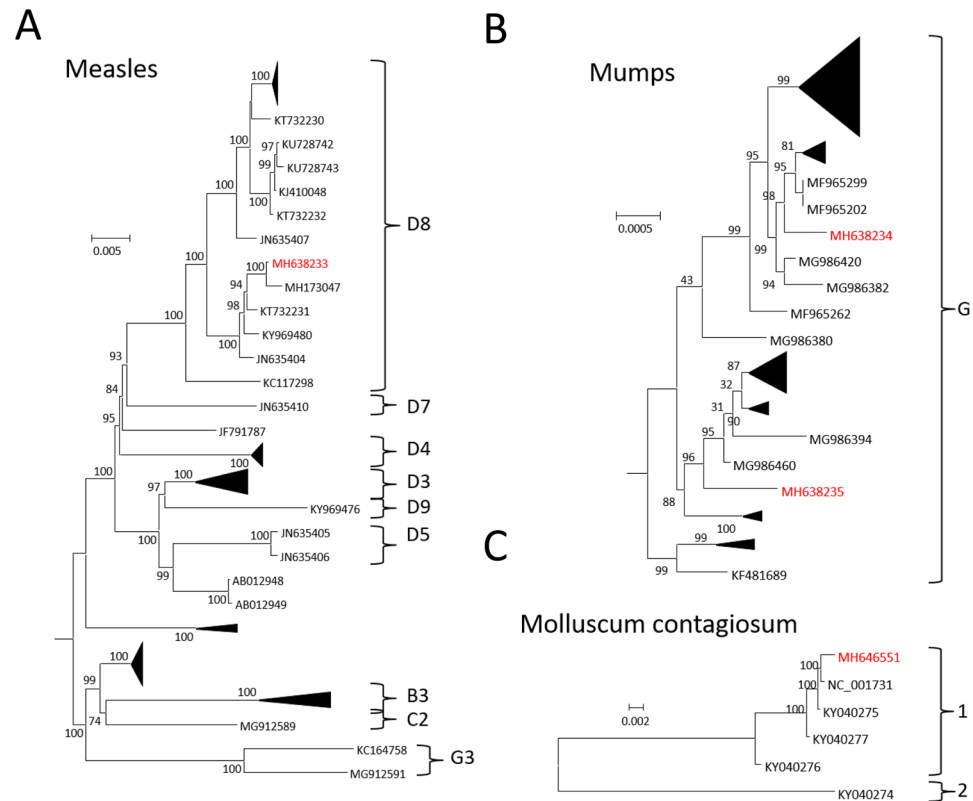


Figure 4. Virus genotyping using phylogenetic methods. Genome alignments were used to generate phylogenetic trees using FastTree for (A) measles, (B) mumps and (C) molluscum contagiosum viruses. The virus sequences originating in this study are shown as their GenBank accession numbers and are depicted in red. The other viruses are represented as their corresponding accession numbers. Genotypes are given where this information was provided. Branch support is given as a percentage.

Ethical restrictions meant we were unable to test for viruses other than those originally requested, making a false positive determination and assessment of specificity impossible. However, it is expected that with the use of stringent assembly and blast search criteria, detection should be highly specific. Nevertheless, the fact that ethics committee did raise such initial concerns about use of the methodology highlights a potential limitation of regarding future use. For example, will clinicians be keen to adopt this given the potential for incidental findings? In our opinion this will hinge on the importance of the clinical question, whereby the risks posed by ancillary findings are outweighed by the need to accurately diagnose and treat the patient.

As the study was restricted to ~50 samples, there were unavoidable biases introduced with a study of this size, such as different frequencies and species of virus genomes found in different sample types. Due to the relatively small number of DNA virus samples tested, further work needs to be done with this group of viruses to be confident of detection. Hence, there is still a need to expand this work to a much larger study to ensure widespread applicability of the method which includes further optimization with regard to all kinds of samples and viral genome types. Additionally, some aspects of the HTS method could be improved. For example, due the ethical restrictions discussed above, draft genome assembly was limited to reference assembly only. Preferably, a *de novo* assembly such as SPAdes²⁹ should be used to minimize the level of bias and detect additional agents if these were present.

Recent work has highlighted the importance of genome data for identifying and responding to virus emergence in which co-circulating viruses can cause clinically similar manifestations among human hosts^{30,31}. Genomic data is of great value in the early stages of emergence (i.e. spillover events) when the nature of a novel agent is unknown³², to determine movement and spread during the outbreak, and changes to the genome which might affect the ability to detect the virus or develop effective vaccines or other therapeutics. Targeted real-time surveillance of vulnerable populations has been identified as one of the most effective measures to flag when more dramatic public health interventions are required. We believe that widespread application of virus genome sequencing to routine virus diagnostics will greatly facilitate this real-time surveillance strategy. To make this a reality, a number of previously identified challenges will need to be overcome³³. In our hands, RNA libraries were constructed for ~\$US120 including labour, and DNA libraries for ~\$20 less. Turn-around time was 48 h for sample preparation and library construction, and 29 h for Illumina sequencing. These will need to be reduced, and further consideration given to improving sensitivity, reducing contamination, dealing with ethical and quality control and assurance concerns, and appropriate bioinformatic support for ease of clinical interpretation. Additional studies are needed to fully examine the actual utility of such methods in epidemiological investigations during outbreak response. As a first step, this work demonstrates that broader application of HTS in clinical

virology is feasible, and gives examples of its application to sequencing of clinical samples directly to enable epidemiological information to be extracted.

Data Availability

All data generated or analyzed during this study are included in this published article (and its Supplementary Information files).

References

- Brown, J. R., Bharucha, T. & Breuer, J. Encephalitis diagnosis using metagenomics: application of next generation sequencing for undiagnosed cases. *J Infect* **76**, 225–240, <https://doi.org/10.1016/j.jinf.2017.12.014> (2018).
- Chan, B. K., Wilson, T., Fischer, K. F. & Kriesel, J. D. Deep sequencing to identify the causes of viral encephalitis. *PLoS One* **9**, e93993, <https://doi.org/10.1371/journal.pone.0093993> (2014).
- Chiu, C. Y. Viral pathogen discovery. *Curr Opin Microbiol* **16**, 468–478, <https://doi.org/10.1016/j.mib.2013.05.001> (2013).
- Chiu, C. Y. *et al.* Diagnosis of Fatal Human Case of St. Louis Encephalitis Virus Infection by Metagenomic Sequencing, California, 2016. *Emerg Infect Dis* **23**, 1964–1968, <https://doi.org/10.3201/eid2310.161986> (2017).
- Chiu, C. Y. *et al.* Identification of cardioviruses related to Theiler's murine encephalomyelitis virus in human infections. *Proc Natl Acad Sci USA* **105**, 14124–14129, <https://doi.org/10.1073/pnas.0805968105> (2008).
- Fremont, M. L. *et al.* Next-Generation Sequencing for Diagnosis and Tailored Therapy: A Case Report of Astrovirus-Associated Progressive Encephalitis. *J Pediatric Infect Dis Soc* **4**, e53–57, <https://doi.org/10.1093/jpids/piv040> (2015).
- Lysholm, F. *et al.* Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. *PLoS One* **7**, e30875, <https://doi.org/10.1371/journal.pone.0030875> PONE-D-11-17038 (2012).
- Naccache, S. N. *et al.* Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. *Clin Infect Dis* **60**, 919–923, <https://doi.org/10.1093/cid/ciu912> (2015).
- Smits, S. L. *et al.* Calicivirus from novel Recovirus genogroup in human diarrhea, Bangladesh. *Emerg Infect Dis* **18**, 1192–1195, <https://doi.org/10.3201/eid1807.120344> (2012).
- Stremlau, M. H. *et al.* Discovery of novel rhabdoviruses in the blood of healthy individuals from West Africa. *PLoS Negl Trop Dis* **9**, e0003631, <https://doi.org/10.1371/journal.pntd.0003631> (2015).
- Yozwiak, N. L. *et al.* Virus identification in unknown tropical febrile illness cases using deep sequencing. *PLoS Negl Trop Dis* **6**, e1485, <https://doi.org/10.1371/journal.pntd.0001485> PNTD-D-11-00551 (2012).
- Fischer, N. *et al.* Evaluation of Unbiased Next-Generation Sequencing of RNA (RNA-seq) as a Diagnostic Method in Influenza Virus-Positive Respiratory Samples. *J Clin Microbiol* **53**, 2238–2250, <https://doi.org/10.1128/JCM.02495-14> (2015).
- Graf, E. H. *et al.* Unbiased Detection of Respiratory Viruses by Use of RNA Sequencing-Based Metagenomics: a Systematic Comparison to a Commercial PCR Panel. *J Clin Microbiol* **54**, 1000–1007, <https://doi.org/10.1128/JCM.03060-15> (2016).
- Greninger, A. L. *et al.* A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS One* **5**, e13381, <https://doi.org/10.1371/journal.pone.0013381> (2010).
- Prachayangprecha, S. *et al.* Exploring the potential of next-generation sequencing in detection of respiratory viruses. *J Clin Microbiol* **52**, 3722–3730, <https://doi.org/10.1128/JCM.01641-14> (2014).
- Thorburn, F. *et al.* The use of next generation sequencing in the diagnosis and typing of respiratory infections. *J Clin Virol* **69**, 96–100, <https://doi.org/10.1016/j.jcv.2015.06.082> (2015).
- Nguyen, A. T. *et al.* Development and evaluation of a non-ribosomal random PCR and next-generation sequencing based assay for detection and sequencing of hand, foot and mouth disease pathogens. *Viral J* **13**, 125, <https://doi.org/10.1186/s12985-016-0580-9> (2016).
- Li, L. *et al.* Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. *J Virol Methods* **213**, 139–146, <https://doi.org/10.1016/j.jviromet.2014.12.002> (2015).
- van den Hurk, A. F., Hall-Mendelin, S., Pyke, A. T., Smith, G. A. & Mackenzie, J. S. Vector competence of Australian mosquitoes for chikungunya virus. *Vector Borne Zoonotic Dis* **10**, 489–495, <https://doi.org/10.1089/vbz.2009.0106> (2010).
- Huang, B., Pyke, A. T., McMahon, J. & Warrilow, D. Complete Coding Sequence of a Case of Chikungunya Virus Imported into Australia. *Genome Announc* **5**, <https://doi.org/10.1128/genomeA.00310-17> (2017).
- Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649, <https://doi.org/10.1093/bioinformatics/bts199> (2012).
- Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**, 1394–1403, <https://doi.org/10.1101/gr.2289704> (2004).
- Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780, <https://doi.org/10.1093/molbev/mst010> (2013).
- Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490, <https://doi.org/10.1371/journal.pone.0009490> (2010).
- Lun, J. H. *et al.* Emerging recombinant noroviruses identified by clinical and waste water screening. *Emerg Microbes Infect* **7**, 50, <https://doi.org/10.1038/s41426-018-0047-8> (2018).
- Rosewell, A., Reinten-Reynolds, T. & Spokes, P. J. EpiReview: Measles in NSW, 2002–2011. *N S W Public Health Bull* **23**, 201–207, <https://doi.org/10.1071/NB12085> (2012).
- Sabbe, M. & Vandermeulen, C. The resurgence of mumps and pertussis. *Hum Vaccin Immunother* **12**, 955–959, <https://doi.org/10.1080/21645515.2015.1113357> (2016).
- Lopez-Bueno, A., Parras-Molto, M., Lopez-Barrantes, O., Belda, S. & Alejo, A. Recombination events and variability among full-length genomes of co-circulating molluscum contagiosum virus subtypes 1 and 2. *J Gen Virol* **98**, 1073–1079, <https://doi.org/10.1099/jgv.0.000759> (2017).
- Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455–477, <https://doi.org/10.1089/cmb.2012.0021> (2012).
- Grubaugh, N. D., Faria, N. R., Andersen, K. G. & Pybus, O. G. Genomic Insights into Zika Virus Emergence and Spread. *Cell* **172**, 1160–1162, <https://doi.org/10.1016/j.cell.2018.02.027> (2018).
- Holmes, E. C., Rambaut, A. & Andersen, K. G. Pandemics: spend on surveillance, not prediction. *Nature* **558**, 180–182, <https://doi.org/10.1038/d41586-018-05373-w> (2018).
- Woolhouse, M. E., Brierley, L., McCaffery, C. & Lycett, S. Assessing the Epidemic Potential of RNA and DNA Viruses. *Emerg Infect Dis* **22**, 2037–2044, <https://doi.org/10.3201/eid2212.160123> (2016).
- Allcock, R. J. N., Jennison, A. V. & Warrilow, D. Towards a Universal Molecular Microbiological Test. *J Clin Microbiol* **55**, 3175–3182, <https://doi.org/10.1128/JCM.01155-17> (2017).

Acknowledgements

This work was funded by Pathology Queensland in 2016, grant number 5488. Many thanks to Assoc. Prof. Ian Mackay for assistance with methods. Thanks also to Carmel Taylor and Dr Alyssa Pyke for helpful comments on the manuscript.

Author Contributions

D.Wa., A.J., D.Wh., R.G. conceived and designed the study. B.H., J.M, G.H., A.DeJ. performed the experiments. D.Wa. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-41830-w>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019